Geography Working Paper Series
No. 2000-4

Introducing the Dissemination Area for the 2001 Census

by

Henry Puderer

Geography Division
Statistics Canada

March 2000

Catalogue no. 92F0138MIE, no. 2000-4

ISSN 1481-174X
ISBN 0-660-18088-X

Aussi disponible en français.

This paper represents the views of the authors and does not necessarily reflect the opinions of Statistics Canada. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by Statistics Canada.

100-07226/

Geography Working Paper Series

The Geography Working Paper Series is intended to stimulate discussion on a variety of topics covering conceptual, methodological or technical work to support the development and dissemination of the division's data, products and services. Readers of the series are encouraged to contact the Geography Division with comments, criticisms and suggestions. A list of titles appears at the end of this paper.

A paper version, catalogue no. 92F0138MPE, is available for $10.00 per issue. Outside Canada the cost is US $10.00. All prices exclude sales tax. There is no charge for downloading the document yourself on the Internet (*http://www.statcan.ca*).

For inquiries about the working paper series, please contact:

GEO-Help
Geography Division
Statistics Canada
Jean Talon Building, 3rd floor
Ottawa, Ontario K1A 0T6

Telephone: (613) 951-3889
FAX: (613) 951-0569
Internet: geohelp@statcan.ca

TABLE OF CONTENTS

ABSTRACT ii

1. INTRODUCTION 1
2. DISSEMINATION AREA DESIGN CRITERIA 1
   - 2.1 Temporal Stability 2
   - 2.2 Minimum Population (Reduced Area Suppression) 3
   - 2.3 Maximum Population (Uniformity) 4
   - 2.4 Intuitive (Visible) Boundaries 4
   - 2.5 Compact Shape 5
   - 2.6 Homogeneity 5
3. OPERATIONAL FACTORS 5
   - 3.1 Dissemination Area Code Structure 5
   - 3.2 Maximum Two-digit Block Code 6
   - 3.3 Production Time-line 6
4. FEEDBACK 6

ACKNOWLEDGEMENTS 6
REFERENCES 7

Appendix 1 - The Block 8
- Map 1. Blocks in an urban area for the 2001 Census 9
- Map 2. Blocks in a rural area for the 2001 Census 9

ABSTRACT

Up to and including the 1996 Census, the enumeration area (EA) served as both the primary collection area and the basic dissemination area. Optimising both the collection and dissemination functions was a difficult undertaking and compromises were made. For the 2001 Census, the new digital cartographic file called the National Geographic Base (NGB) and the block program have made it possible to create separate enumeration areas for data collection and dissemination areas for data dissemination. This working paper describes the design criteria that have been specified for the implementation of the dissemination area (DA). These criteria address the improvements most frequently requested by users such as increased temporal stability, reduced area suppression, intuitive boundaries, compactness and homogeneity. In addition to integrating and reflecting the recommendations from the user community, the dissemination area design incorporates operational factors related to the DA code structure, maximum code size and production timeframes.

1. INTRODUCTION

Up to and including the 1996 Census, the **enumeration area** (EA) served as both the primary collection area and the basic dissemination area. Optimising both the collection and dissemination functions was a difficult undertaking and compromise eroded the optimum for both. As a consequence, there have been long-standing requests from the user community to make modifications to the enumeration area that would improve it as a dissemination area (Statistics Canada, 1994, 1999). The most frequently requested improvements were for increased temporal stability, reduced area suppression, intuitive boundaries, compactness and homogeneity. It was never possible to make these modifications to the EA without seriously compromising its function as a collection unit. Nor was it possible to produce a separate dissemination area at a cost that was not prohibitive. But the situation has changed as we prepare for the 2001 Census.

First, a **National Geographic Base** (NGB), which is a digital cartographic base for all of Canada, has been created.[1] The availability of a digital base means that software can be developed to economically automate the delineation of a dissemination area designed to address the above noted users' requests.

Second, the 2001 Census will see the introduction of the **block** program. Essentially, blocks are the polygons formed by the intersection of streets, although these morphological blocks are split when the boundaries of selected standard geographic areas (like municipalities) or collection requirements dictate. At the time of enumeration, all dwellings will be geographically referenced to specific blocks, created prior to Census day. The availability of the block is critical. The design of the collection geography can be largely separate from the dissemination geography because these blocks can be aggregated one way for collection and another way for dissemination.[2]

Because it is now both technically and economically possible, the decision has been made to create the **dissemination area (DA)**, a standard geographic area designed specifically for data output, to replace the enumeration area for dissemination. Dissemination area delineation will be accomplished using a system developed in-house called Generalized Area Delineation System (GArDS). GArDS aggregates small geographic areas (in this case, blocks) according to a set of delineating or design criteria to produce a set of the desired geographic areas (in this case, dissemination areas) in a larger geographic work area called the largest delineation zone (LDZ). The design criteria are assigned penalty weights. The solution that is accepted is the one that has the lowest total penalty weight, which is an aggregate of the penalty weights for all of the combined criteria for all geographic areas within the LDZ. The rate at which the penalty weight increases can also be controlled so the impact of the different criteria can be modified according to their assessed importance to the desired solution. This working paper describes the design criteria that have been specified for the implementation of the dissemination area.

2. DISSEMINATION AREA DESIGN CRITERIA

On the basis of feedback from the user community, six criteria were considered in the design of the DA. In order of priority, these are:

- Temporal stability
- Reduced area suppression
- Uniformity
- Intuitive boundaries
- Compact shape
- Homogeneity.

---

[1] The National Geographic Base (NGB) is a digital base being created in partnership with Elections Canada. Source files for the NGB include the 1996 Census Street Network Files (SNF) produced by the Geography Division, Statistics Canada; Natural Resources Canada's digital coverage (NTDB) and the Digital Chart of the World (DCW) coverage for the northern parts of Canada, plus digital files from Elections Canada (NTDB with names tagged to streets).

[2] The block is also a new geographic area for dissemination of the 2001 Census. Current plans envision releasing population and dwelling counts by block but no characteristic data. The smallest geographic area for which characteristic data will be released is the dissemination area. For more details about the block program see Appendix 1.

This does not mean that the DA will be perfect in all respects because of the many trade-off conditions among the design criteria. However, what can be realised is an improvement in the characteristics of the DA relative to those of the EA, which in turn will improve the usability of the census data that are subsequently made available. What follows is a discussion of the rationale for the individual criteria and the trade-off process that was an inherent part of the selection.

2.1 Temporal Stability

The boundaries of enumeration areas change from one census to another because of changes to the boundaries of standard geographic areas, the amount of population growth and changes to field collection specifications.[3] The degree of change varies from census to census.

Users often see these changes as a frustrating inconvenience. They can confound longitudinal analysis completely or force difficult and sometimes costly adjustments to the area of study in order to get comparable data across censuses.

Because instability in the EA was in large measure a consequence of the requirement for the EA to respect changes to the boundaries of standard geographic areas, it would be desirable, from a temporal stability perspective, to have the DA totally independent of any changes to boundaries of standard areas. This option could be considered because the block could now provide the link between data collection areas and the standard geographic areas for data dissemination.

However, user feedback had indicated that for maximum utility, dissemination areas should not straddle municipal (CSD) boundaries. A DA that respects the boundaries of CSDs would be especially useful where there is no census tract (CT) program. A DA that respects the boundaries of census tracts increases the utility of the CT program by providing the capability to drill down further to analyse intra-CT distributions. Therefore, after considering various alternatives that represented trade-offs between DA stability and utility, the approach chosen will require the DA to respect the boundaries of census subdivisions (CSDs) and census tracts (CTs).

What is the impact of this approach on DA stability? Census tract boundary revisions rarely occur and only when essential. Therefore, respecting the boundaries of CTs has minimal impact on the stability of the DA. However, future CT splits and CT creation or changes to CT boundaries – usually on the periphery of a census metropolitan area (CMA) or census agglomeration (CA) – will result in DA changes. However, these changes are localised and once put into effect, stability returns to the local area as the zone of change pushes outward. Similarly, change is anticipated where a CSD is incorporated or where a CSD boundary changes as a consequence of an annexation. Again, the expectation is that the change will be localised and a one-time change. In general, where growth occurs, DA splits are likely, as in the case of the CT program. If decline should occur, aggregations could be considered, especially with respect to the minimum population threshold criterion. Minor boundary changes to either a CT or a CSD would in most instances result in similar changes to the affected DA, whereas significant changes (for example, a census agglomeration entering the census tract program) could result in a complete redo for the area affected.

There are two aspects to the temporal stability criterion. The above has discussed the forward part – building a solid base for the DA starting with the 2001 Census and maximising stability for the 2006 Census and beyond. But what about the reverse aspect, namely, retaining comparability to the 1996 Census EA? During consultations, opinion on this particular item was more divided than on any of the other criteria. One group clearly wished to see comparability to the 1996 Census EA as a design criterion while another group favoured dropping this historic baggage. In the end, the latter view prevailed because

[3] In order to aggregate census data to the standard geographic areas for dissemination, the EA must respect the boundaries of the standard geographic areas. These boundaries are subject to change from census to census. For example, between censuses, municipal boundaries change as municipalities are created or dissolved or annex parts of adjacent municipalities (municipalities are a major component of the census subdivision (CSD) standard geographic area). Historically we could expect 20% of the 6,000 CSDs to change boundaries from one census to another. The federal electoral district (FED) is another standard geographic area that EAs must respect. When new FED Representation Orders come into effect, 20% of the previous census EAs are altered. Population growth and the associated dwelling growth, changes to the road network, and changes to field collection specifications, such as the maximum number of dwellings per EA, also force changes to the EA from one census to the next.

preserving comparability also meant preserving sub-optimal DAs. It would also constrain the creation of new DAs, thereby creating an environment supporting the generation of more sub-optimal DAs.

2.2 Minimum Population (Reduced Area Suppression)

The criteria for delineating enumeration areas for census collection conflict with the techniques used to protect the confidentiality of individual responses for dissemination. One technique called area suppression results in the deletion of all census characteristic data for geographic areas with population counts fewer than 40 (fewer than 250 in the case of income data). [4]

The application of area suppression rules to the 49,361 enumeration areas (EAs) for the 1996 Census meant that income data were suppressed for 13,085 EAs (27%) with fewer than 250 people, while other characteristic data were suppressed for 4,765 EAs (10%) with fewer than 40 people. In fact, 2,278 (5%) of the 49,361 EAs in 1996 had zero population. The need to reduce such high levels of area suppression for the smallest standard geographic area/dissemination unit is axiomatic. In the past it was not attainable given the dual roles of the enumeration area. However, it is largely attainable as part of the dissemination area design.

A minimum population of 500 has been specified as the target count for the dissemination area. A penalty weight is applied to the automated solution if the population count for the DA falls below this target value. The value of the weight increases as the difference between the solution and the target count increases. The rate at which the penalty weight increases can also be controlled so the impact of the different criteria can be modified according to their assessed importance to the DA solution. The minimum population criterion has been assigned the highest penalty weights. Nonetheless, DAs with population counts below 500 will still be delineated. The DA solution that is accepted is the one that has the lowest total penalty weight, which is an aggregate of the penalty weights for all of the combined criteria for all DAs within the largest delineation zone (LDZ). [5, 6] However, given that the minimum population criterion has the largest penalty weight, the DA in most instances will not have population counts significantly below the target value and certainly DAs below 250 will be rare, subject to two caveats.

The first caveat is that DAs are required to respect the boundaries of CSDs and CTs (because of the "temporal stability" criterion described above). Therefore, the automated system will produce solutions for largest delineation zones defined by the intersection of CSDs and CTs. There will be a number of individual CSDs and CTs that fall below the target population. For instance, using 1996 Census results, there were 2,329 CSDs and 94 CTs with a population less than 500. The number of CSDs and CTs with fewer than 250 people is 1,449 and 69 respectively and the number with below 40 people are 637 and 40 respectively. As well, there are CTs that straddle CSDs, which will result in LDZs with population counts below the target 500. There are 256 CTs that straddle CSD boundaries. Assume that something like 50,000 DAs will be created for the 2001 Census and that all of the 256 straddling CTs create LDZs of fewer than 40 people. Then we can expect that at least 6% of the DAs will have fewer than 500 population, 4% will have fewer than 250 and 2% will have fewer than 40. Although not perfect, it will be a significant improvement compared to enumeration areas.

The second caveat relates to the source of the population data that will be used in the DA delineation. We would like to use the 2001 Census population counts by block. However, the date when the final set of DAs is required may preclude use of the 2001 Census data. This means the DA delineation will have to

---

[4] The other technique used is random rounding. Under this method all figures, including totals, that are greater than 10 are randomly rounded either up or down to a multiple of "5" and figures that are 10 or less are randomly rounded to "0" or "10".

[5] The LDZ is the geographic work area for which a solution is being sought. For the DA, it is defined by the intersection of census subdivisions (CSDs) and census tracts (CTs).

[6] This approach does not directly take into account the spread of the individual scores for each DA within the LDZ. As a consequence, use of the overall solution with the lowest aggregate penalty score within the LDZ could result in a poor DA while generating optimal solutions for the other DAs within the LDZ. An alternative approach could ensure that individual DA scores are considered so that the creation of a poor DA is prevented, but this would cause slightly poorer quality solutions for the other DAs within the LDZ, which means the lowest aggregate penalty score solution would not be selected. Whether the second approach is better than the current one being programmed needs to be assessed during the testing phase.

use population estimates by block. The methodology for this aspect of the DA creation has not been investigated sufficiently to describe either the process or the anticipated quality. Nonetheless, we envision distributing 1996 EA counts across the corresponding blocks with some adjustment for growth. As a consequence of the associated error with this process, we expect cases where the DA population count is lower than the target 500. However we also expect the estimates to be sufficiently robust to prevent the creation of DAs with a population count below 250.

2.3 Maximum Population (Uniformity)

The purpose of this criterion is to keep the DA population size uniform. This criterion will have the benefit of generating more DAs within the LDZ. This will help when trying to fit DAs to other geographic areas of interest. Without an explicit maximum population criterion, the delineation would tend more towards larger and fewer DAs within the LDZ.

The maximum population target is also 500. It is not as critical as the minimum criterion since it does not affect the availability of data. Therefore, the penalty scores assigned to the maximum population target are less than those assigned to the minimum target. Consequently, the overall tendency will be to produce DAs greater than the target of 500, but to constrain the solutions so that DAs that significantly exceed the target are not created. As with the minimum population criterion, the use of population estimates will affect the final results but this is not considered too serious in this instance.[7]

During the deliberations concerning the minimum and maximum population criteria, a related issue was brought into discussion. The issue was of the method of selecting population targets and data reliability considerations. In general, because data reliability decreases with cell size, the objective is to try to have cell sizes above some minimum size. Consequently, from the perspective of data reliability, the larger the minimum threshold, the better. However, increasing the minimum threshold beyond a certain level runs counter to the objectives of a basic standard geographic dissemination area, so some maximum limit is required. Table design is also a critical component when addressing data reliability. In the end, trying to generate uniformly sized DAs (in terms of population size) was seen as the goal since the resulting consistency would benefit in the design of optimum output tables. The target of 500 was seen as a good all round choice with respect to both area suppression and data reliability concerns.

2.4 Intuitive (Visible) Boundaries

Enumeration areas were required to respect all the boundaries of the standard geographic areas in order to aggregate the data to these areas for dissemination. On occasion these standard geographic areas had "invisible" boundaries, i.e., some or all of the boundary did not follow any visible feature on the ground (such as a street, railway, water feature, power transmission line, etc.). As a consequence, the EA boundary also had an invisible component. Furthermore, in the absence of suitable visible features, invisible EA boundaries were sometimes deliberately created in order to limit the size of the EA to reflect collection workload specifications. EA boundaries that follow invisible features can make it difficult for users to determine the geographic coverage of an EA and the associated uncertainty can raise questions about the quality of the data application. Consequently, the user community has requested that only visible and easily identifiable boundaries be used to delineate standard geographic areas – what generically have been called "intuitive" boundaries.

Because DAs are aggregations of blocks, the intuitive boundary criterion is in most instances assured because blocks respect roads. However, blocks are also required to follow the boundaries of selected standard geographic areas and any additional limits required by the collection operation to constrain workloads. As described above, some of these boundaries (or at least sections) are invisible. Therefore, the task for the DA creation is to apply significant penalty scores to block boundaries that are invisible. Again, as is the case with the other criteria, this does not prevent the creation of DAs with invisible boundaries (or at least partially invisible boundaries), but it significantly limits the likelihood of such outcomes. The exception condition to the preceding statement is the fact that the DA respects the boundaries of all census subdivisions (CSDs) and census tracts (CTs) and as a consequence will have invisible boundaries to the extent that the CSDs and CTs do.

---

[7] Low population DAs will result where the distribution of the 1996 EA population by block is inaccurate and high population DAs will result where the growth adjustment is not accurate.

2.5 Compact Shape

As with intuitive boundaries, the demands placed on the EA did result in the generation of strange EAs in terms of shape. For example, if two standard geographic areas straddled each other producing a long narrow sliver where the two boundaries overlapped, an EA needed to be created. Similarly, strange EA shapes resulted where accessibility was a problem. Oddly shaped EAs can create problems when users are attempting to create a "best fit" to their area of interest via the aggregation of EAs. Therefore, users have identified the need to have the dissemination areas as compact as possible.

In order to address this requirement, a compactness index is being included in the DA design. The index is derived by comparing the area of the DA to its perimeter. A square has an index value of one (1.0) while a polygon that is twice as wide as it is long will have a compactness index of two (2.0). In general, if a polygon is stretched so that one side is "n" times longer than the other side, the compactness index will be equal to "n". Values greater than one (1.0) reflect increasingly less compactness as the DA shape differs more and more from that of a perfect square.

Again, the inclusion of this criterion does not prevent what could be judged as inappropriate results. The inclusion of this criterion only makes certain types of DA delineations more probable.

2.6 Homogeneity

On the basis of submissions from the user community, homogeneity is the last important factor suggested for inclusion in the delineation of a standard dissemination area. Dwelling type is most often suggested as the variable that should be used to achieve the desired homogeneity. However, of the six key factors recommended for inclusion in the dissemination area design, it is the only one that we have elected not to explicitly include in the design.

The reason for this decision is the lack of the appropriate data by block. Support of this criterion would require that estimates of dwelling counts, by type of dwelling, be generated by block from the 1996 EA level census data. It was concluded that the quality of dwelling type estimates that could be generated would simply be inadequate to the task, so inclusion of an explicit homogeneity criterion was dropped.[8]

3. OPERATIONAL FACTORS

In addition to the recommendations from the user community, there are also operational factors that need to be included in the dissemination area design. The most important of these is the structure of the DA code.

3.1 Dissemination Area Code Structure

Early in the design process the need was identified to have a hierarchical structure represented by an eight-digit code with embedded geographic proximity. These operational requirements were determined or previously satisfied by the province / federal electoral district / enumeration area (PR / FED / EA) structure. The code size (eight-digit) is a consequence of the PR / FED / EA code size namely, 2 + 3 + 3 = 8. This code size is required to avoid redesigning the many programs that need to access this code during census processing. The hierarchical structure with geographic proximity is required to support edit and imputation when a geographically close donor record is required.

All code requirements will be accomplished via a province / census division / dissemination area (PR / CD / DA) hierarchy. The eight-digit code maximum is satisfied with PR (2) + CD (2) + DA (4). Not only is the hierarchical structure requirement satisfied, the number of units at each level is approximately

[8] When it became clear that the homogeneity criterion could not be supported directly, there was some thought that homogeneity could still be supported indirectly by keeping the DA small and compact. In other words, if DAs were comprised of from two to four blocks, then homogeneity would be obtained indirectly since dwelling type often tends to be consistent from block to block without sudden transitions. Although this view is unsubstantiated, it is likely that DAs of two to four blocks will be well below the norm in terms of population size.

the same since there are 301 FEDs and 288 CDs. As with the EA, geographic proximity will be embedded in the code by assigning DA codes in a serpentine manner within census divisions.[9]

3.2 Maximum Two-digit Block Code

A second operational requirement may limit the number of blocks that can be included in a DA to 99. This results from a database design that provides for a two-digit block code. Although this design limitation could be changed (e.g., a three-digit code could be created), it would be costly to implement. Consequently, the decision has been made to wait for the first DA test. The first test will not include this limitation. A final decision will be made after determining how many times DAs will be created exceeding the limit of 99 blocks and assessing the impact this limit would have on the creation of dissemination areas.[10, 11]

3.3 Production Time-line

A three-month elapsed time production window is envisioned for the creation of the 2001 DA structure (from May 2001 to July 2001). Not only is this operational requirement forcing the use of block population estimates for the creation of DAs, but it may also further influence the DA specifications if the required production time exceeds that available. However, there are a number of potential solutions to be investigated before the specifications would need to be reduced to obtain the required time reduction. First, increased computing power is a possible solution. Streamlining the algorithms without reducing the specification requirements is also a possible solution. Limiting the number of solutions to be considered by limiting the amount of computational time for a LDZ is possible, although this approach does not ensure the best solution will be found. It is important to point out however that discussion of this requirement is highly speculative and without empirical production experience at this time. Nonetheless, it is an operational requirement with the potential to affect DA specifications and as such is included in this paper for completeness.

4. FEEDBACK

The specifications for generating the dissemination area for the 2001 Census are for the most part fixed. Nevertheless, we are always interested in receiving feedback. Comments, suggestions or inquires about any aspect covered in this working paper are most welcome and should be directed to GEO-Help in Geography Division.

ACKNOWLEDGEMENTS

This working paper is a compilation of ideas and research conducted over the past five years and reflects the efforts of a number of staff from the Geography Division (Pessoa, Earl and Puderer, 1996). The contributions of Louise Earl, Peter Murphy, José Pessoa and the reviews of Carolyn Weiss and the Dissemination Area Working Team chaired by Rick Rideout are acknowledged. Thanks are extended to Karole Kidd who edited the final draft of this working paper and Réjeanne Loranger who reviewed the French translation.

---

[9] This code structure is an internal operational requirement. Alternative structures for dissemination purposes are possible and can be considered; i.e., suggestions are welcome.

[10] Given the maximum population target of 500, we believe the number of times an optimum DA would exceed 99 blocks will be low. In other words, we do not believe this limit on the number of blocks in a DA will adversely affect their formation.

[11] The limit of 99 blocks also applies to the number of blocks per enumeration area. Therefore, some collapsing of blocks will result but the incidence will be low.

REFERENCES

Pessoa, José, Earl, Louise and Puderer, Henry (1996). *The 2001 Logical Dissemination Unit Model: A Proposal for the Development of a New Small Area Geographic Unit for the Dissemination of Data*, unpublished report, Geography Division, Statistics Canada (Ottawa).

Statistics Canada (1994). *1996 Census Consultation Report*, 1996 Census of Canada, Report No. 2 (Ottawa: Statistics Canada).

Statistics Canada (1999). *2001 Census Consultation Geography Report*, 2001 Census of Canada, Catalogue No. 92-126-SPB (Ottawa: Industry Canada).

Appendix 1

THE BLOCK

Up to the 1996 Census, dwellings and their occupants (i.e., census data) were geographically referenced to the enumeration area at the time of collection. Starting with the 2001 Census, the data will be geographically referenced to the block at the time of collection.[12]

First and foremost, blocks are the polygons formed by the intersection of streets[13] (see Maps 1 and 2). These natural or morphological blocks will be split to form two blocks wherever a block is traversed by the boundaries of selected standard geographic areas, namely federal electoral districts (FED), census subdivisions (CSD), census tracts (CT) or designated places (DPL). This is required in order to be able to aggregate the data to all standard geographic areas for dissemination.

A block may also be formed to satisfy collection requirements. For example, workload limits require that blocks be formed for large apartment buildings or collective dwellings. In these cases, blocks within blocks will result. As well, where the road network is sparse or even non-existent, enumeration area boundaries may be established to constrain the coverage area for the census enumerator. In these situations, the block will respect the boundary of the enumeration area.

Blocks are aggregated to build enumeration areas and dissemination areas. Since blocks are primarily an artifact of the road network, the number of blocks and their creation is a function of how up-to-date the road network database will be prior to the census. It is not possible to have a road network reflecting exactly the situation on census day. This will be especially true in high growth areas.[14]

Population and dwelling counts will be disseminated by block. Characteristic data will not be available by block. The smallest standard geographic area for which characteristic data will be provided will be the DA. However, the block will replace the EA outside block-face coverage as the basic unit for constructing user-defined areas, thus providing increased precision in the delineation of these areas.

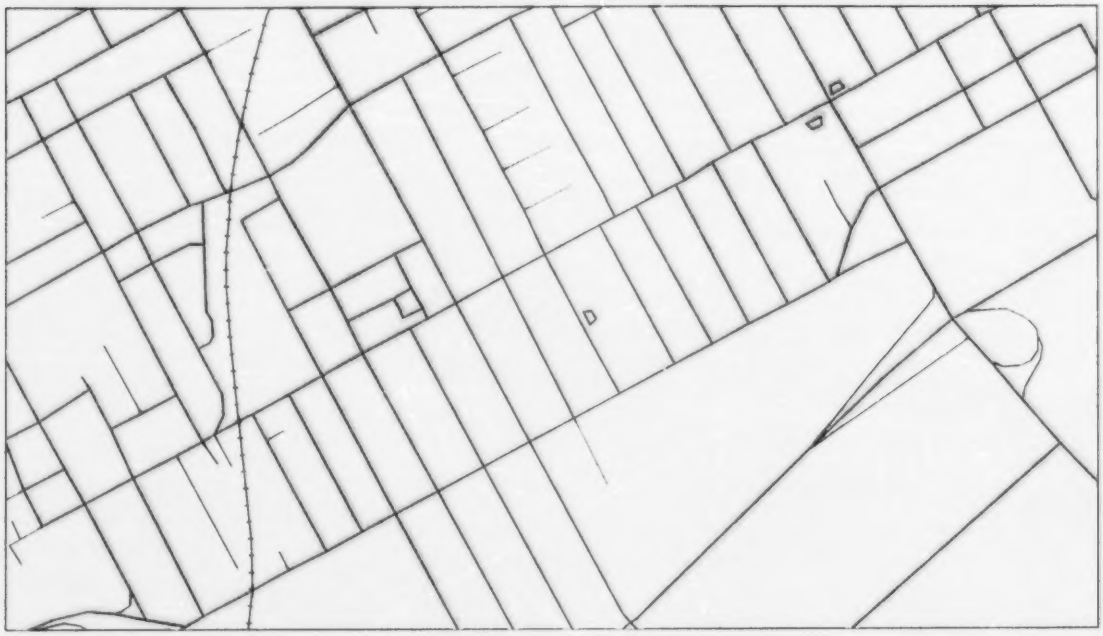
---

[12] Census data have been and will continue to be referenced to the block-face to support data retrieval for user defined areas. But this linkage to the block-face is done as a post-collection activity and only for selected areas.

[13] Certain morphological blocks (e.g., road medians, ramp areas) will be merged with adjacent blocks.

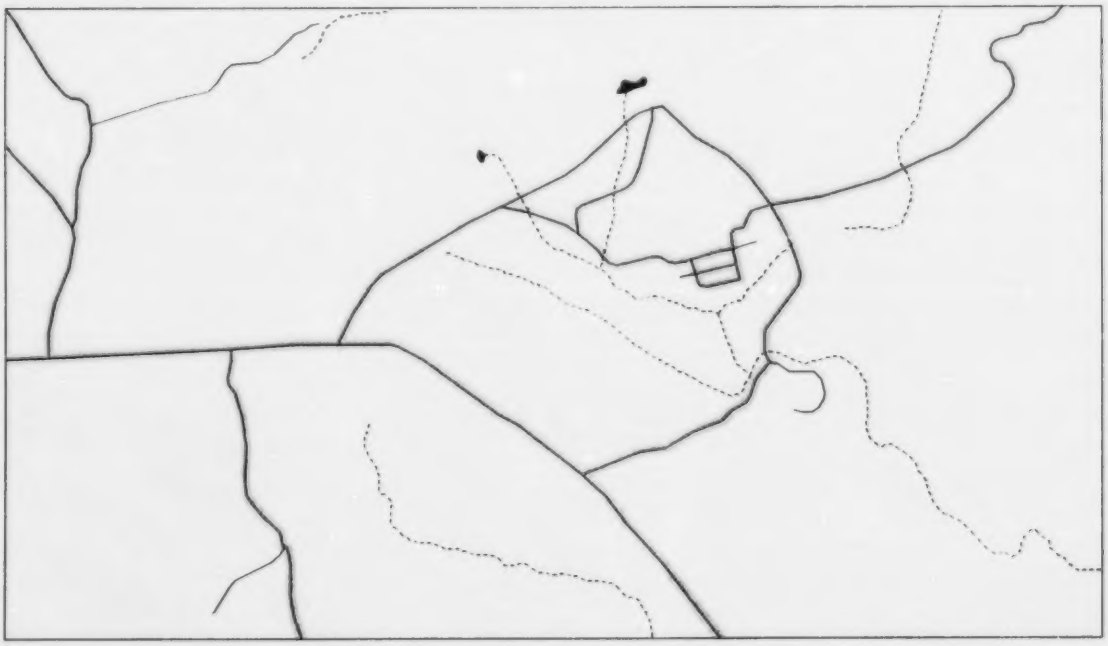
[14] At the writing of this working paper, the road network on the NGB was up-to-date to about 1995 and generated about 630,000 blocks. However, the road network continues to be updated. The latest round of updates currently in progress will bring the base up-to-date as of spring 1997.

Map 1. Blocks in an urban area for the 2001 Census



Blocks
Roads
Railroads
Lakes / Ponds
Rivers / Creeks

Map 2. Blocks in a rural area for the 2001 Census



Geography Working Paper Series

Number

1993-1 ***A Comparison of Census Geographic Areas of Canada and the United States***, Carolyn Weiss, Michael Ratcliffe and Nancy Torrieri (November 1993)

1996-1 ***Whither Land Area? A Case for Retaining the Status Quo for the 1996 Census***, Carolyn Weiss (December 1996)

1998-1 ***Automated Land Area Tests for the 2001 Census: Preliminary Results Using the 1996 Digital Cartographic Files***, Carolyn Weiss and Augustine Akuoko-Asibey (April 1998)

2000-1 ***Census Metropolitan Area and Census Agglomeration Influenced Zones (MIZ) with Census Data***, Sheila Rambeau and Kathleen Todd (January 2000)

2000-2 ***Census Metropolitan Area and Census Agglomeration Influence Zones (MIZ): A Description of the Methodology***, Chuck McNiven, Henry Puderer and Darryl Janes (January 2000)

2000-3 ***Delineation of Canada's North: An Examination of the North–South Relationship in Canada***, Chuck McNiven and Henry Puderer (January 2000)

2000-4 ***Introducing the Dissemination Area for the 2001 Census***, Henry Puderer (March 2000)

Série de documents de travail de la géographie
N° 2000-4

Présentation de la région de diffusion pour le Recensement de 2001

par

Henry Puderer

Division de la géographie
Statistique Canada

Mars 2000

N° 92F0138MIF au catalogue, n° 2000-4

ISSN 1481-1758
ISBN 0-660-96311-6

Also available in English.

Ce document représente les opinions des auteurs et ne reflète pas nécessairement le point de vue de Statistique Canada. Toute mention d'une marque, d'un produit ou d'une entreprise sert à des besoins représentatifs seulement et ne représente pas l'appui de Statistique Canada.

Série de documents de travail de la géographie

La Série de documents de travail de la géographie vise à stimuler des discussions sur une variété de sujets reliés au travail conceptuel, méthodologique et technique à l'appui de l'élaboration et de la diffusion des données, des produits et des services de la division. Nous encourageons les lecteurs de la série à communiquer avec la Division de la géographie pour lui fournir leurs commentaires, critiques et suggestions. Une liste des titres apparaît à la fin du document.

Une version papier, n° 92F0138MPF au catalogue, est disponible pour 10,00 $ par numéro. À l'extérieur du Canada, le coût est de 10,00 $ US par numéro. Les prix n'incluent pas la taxe de vente. Il n'y a pas de frais pour télécharger le document vous même à partir du site Internet (*http://www.statcan.ca*).

Pour toute demande de renseignements au sujet de la série de documents de travail, veuillez communiquer avec :

Info-GÉO
Division de la géographie
Statistique Canada
Édifice Jean-Talon, 3e étage
Ottawa (Ontario) K1A 0T6

Téléphone : (613) 951-3889
Télécopieur : (613) 951-0569
Internet: geohelp@statcan.ca

TABLE DES MATIÈRES

RÉSUMÉ ii

1. INTRODUCTION 1
2. CRITÈRES DE CONCEPTION RELATIFS AUX RÉGIONS DE DIFFUSION 2
   - 2.1 Stabilité temporelle 2
   - 2.2 Population minimale (réduction de la suppression de données sur des régions) 3
   - 2.3 Population maximale (uniformité) 4
   - 2.4 Limites intuitives (visibles) 5
   - 2.5 Forme compacte 6
   - 2.6 Homogénéité 6
3. FACTEURS OPÉRATIONNELS 6
   - 3.1 Structure du code de la région de diffusion 6
   - 3.2 Code d'îlot à deux chiffres maximum 7
   - 3.3 Calendrier de production 7
4. RÉTROACTION 7

REMERCIEMENTS 7
RÉFÉRENCES 8

Annexe 1 - L'îlot 9
- Carte 1. Îlots dans une région urbaine pour le Recensement de 2001 10
- Carte 2. Îlots dans une région rurale pour le Recensement de 2001 10

RÉSUMÉ

Jusqu'au Recensement de 1996 inclusivement, le secteur de dénombrement (SD) servait à la fois de secteur de collecte principal et de région de diffusion de base. L'optimisation des fonctions de collecte et de diffusion était une tâche difficile et il a fallu faire des compromis. En ce qui concerne le Recensement de 2001, le nouveau fichier cartographique numérique, la Base géographique nationale (BGN), et le programme des îlots ont permis de créer des secteurs de dénombrement distincts aux fins de la collecte de données, ainsi que des régions de diffusion pour la diffusion des données. Le présent document de travail décrit les critères de conception qui ont été déterminés pour la mise en œuvre de la région de diffusion (RD). Ces critères portent sur les améliorations le plus souvent demandées par les utilisateurs, telles que la stabilité temporelle accrue, la réduction de la suppression de données sur des régions, les limites intuitives, la compacité et l'homogénéité. En plus d'intégrer et de refléter les recommandations des utilisateurs, la conception de la région de diffusion incorpore des facteurs opérationnels liés à la structure du code de la région de diffusion, à la taille maximale du code et aux calendriers de production.

1. INTRODUCTION

Jusqu'au Recensement de 1996 inclusivement, le **secteur de dénombrement** (SD) servait à la fois de secteur de collecte principal et de région de diffusion de base. L'optimisation des fonctions de collecte et de diffusion était une tâche difficile et les compromis adoptés ont réduit les possibilités d'optimiser ces deux fonctions. Par conséquent, les utilisateurs demandent depuis longtemps d'apporter des modifications au secteur de dénombrement pour pouvoir mieux s'en servir comme région de diffusion (Statistique Canada, 1994, 1999). Les améliorations le plus souvent demandées consistaient à accroître la stabilité temporelle, à réduire la suppression de données sur des régions, à établir des limites intuitives, et à rehausser la compacité et l'homogénéité. Il n'a jamais été possible d'apporter ces modifications au SD sans compromettre gravement sa fonction d'unité de collecte. Il n'était pas non plus possible de produire une région de diffusion distincte à un coût qui ne serait pas inabordable. La situation a cependant changé maintenant que nous nous préparons au Recensement de 2001.

Pour commencer, la **Base géographique nationale** (BGN), qui est une base cartographique numérique pour l'ensemble du Canada, a été créée[1]. La disponibilité d'une base numérique signifie qu'un logiciel peut être développé afin d'automatiser économiquement la délimitation d'une région de diffusion en vue de répondre aux demandes des utilisateurs mentionnées plus haut.

Ensuite, le Recensement de 2001 verra l'introduction du programme des **îlots**. En gros, les îlots sont les polygones formés par l'intersection des rues, bien que ces îlots morphologiques soient divisés là où les limites des régions géographiques normalisées choisies (comme les municipalités) ou les besoins de collecte l'exigent. Au moment du dénombrement, tous les logements seront géographiquement rattachés à des îlots particuliers, créés avant le jour du Recensement. La disponibilité des îlots est cruciale. La conception de la géographie de collecte peut être séparée de la géographie de diffusion en grande partie parce que ces îlots peuvent être regroupés d'une certaine façon pour la collecte et d'une autre façon pour la diffusion[2].

Étant donné que la possibilité existe maintenant, tant sur le plan technique qu'économique, on a décidé de créer la **région de diffusion (RD)**, une région géographique normalisée conçue spécifiquement pour la production de données, afin de remplacer le secteur de dénombrement aux fins de diffusion. On délimitera la région de diffusion au moyen d'un système développé par Statistique Canada, qui s'appelle Système automatisé de regroupement de territoires (SARTe). Le SARTe regroupe de petits secteurs géographiques (dans ce cas-ci, des îlots) selon un ensemble de critères de délimitation ou de conception en vue de produire un ensemble de régions géographiques voulues (dans ce cas-ci, les régions de diffusion) dans une région de travail géographique plus grande qui s'appelle la plus grande région de délimitation. On attribue des pondérations de pénalité aux critères de conception. La solution acceptée est celle qui a la plus faible pondération de pénalité totale, qui est un agrégat des pondérations de pénalité pour tous les critères combinés de toutes les régions géographiques à l'intérieur de la plus grande région de délimitation. On peut aussi contrôler le taux d'augmentation de la pondération de pénalité, ce qui permet de modifier les répercussions des différents critères en fonction de leur importance évaluée pour la solution voulue. Le présent document de travail décrit les critères de conception qui ont été établis pour la mise en œuvre de la région de diffusion.

---

[1] La Base géographique nationale (BGN) est une base numérique créée en partenariat avec Élections Canada. Les fichiers d'origine pour la BGN comprennent les Fichiers du réseau routier (FRR) du Recensement de 1996 produits par la Division de la géographie à Statistique Canada; la couverture numérique (BNDT) de Ressources naturelles Canada et la couverture de la Carte mondiale numérique (Digital Chart of the World – DCW) pour les parties nordiques du Canada, ainsi que des fichiers numériques d'Élections Canada (la BNDT avec des noms étiquetés aux routes/rues).

[2] L'îlot est aussi la nouvelle région géographique pour la diffusion du Recensement de 2001. Les plans actuels prévoient la diffusion de chiffres de population et de logements par îlot sans données sur les caractéristiques. La plus petite région géographique pour laquelle des données sur les caractéristiques seront diffusées, c'est la région de diffusion. Pour obtenir des détails supplémentaires sur le programme des îlots, se reporter à l'annexe 1.

2. CRITÈRES DE CONCEPTION RELATIFS AUX RÉGIONS DE DIFFUSION

En se fondant sur la rétroaction des utilisateurs, on a étudié six critères pendant la conception des régions de diffusion. Par ordre de priorité, ce sont les critères suivants :

- La stabilité temporelle
- La réduction de la suppression de données sur des régions
- L'uniformité
- Les limites intuitives
- La forme compacte
- L'homogénéité.

Cela ne signifie pas que la région de diffusion sera parfaite à tous les égards en raison des nombreuses conditions de compromis appliquées aux critères de conception. Ce qu'on peut réaliser cependant, c'est une amélioration des caractéristiques de la RD par rapport à ceux du SD qui, à son tour, permettra d'améliorer l'utilisation des données de recensement qui deviendront disponibles par la suite. Ce qui suit est une discussion de la justification des critères individuels et du processus de compromis qui faisait partie intégrante de la sélection.

2.1 Stabilité temporelle

Les limites des secteurs de dénombrement changent d'un recensement à l'autre en raison des changements apportés aux limites des régions géographiques normalisées, du taux de croissance démographique et des modifications apportées aux spécifications de collecte sur le terrain[3]. Le degré de changement varie d'un recensement à l'autre.

Les utilisateurs considèrent souvent que ces changements sont un dérangement énervant. Ces changements peuvent donner entièrement tort à l'analyse longitudinale ou obliger à faire des ajustements difficiles et parfois coûteux au secteur étudié afin d'obtenir des données comparables pour tous les recensements.

Comme l'instabilité dans le SD découle en grande partie de l'exigence voulant que le SD respecte les changements des limites des régions géographiques normalisées, il serait souhaitable, d'un point de vue de stabilité temporelle, que la région de diffusion soit entièrement indépendante de toutes les modifications apportées aux régions normalisées. Cette option pourrait être envisagée parce que l'îlot peut maintenant constituer le lien entre les secteurs de collecte de données et les régions géographiques normalisées pour la diffusion de données.

Toutefois, les utilisateurs ont signalé que, pour maximiser l'utilité, les régions de diffusion ne devraient pas chevaucher les limites municipales (subdivision de recensement). Une RD qui respecte les limites des subdivisions de recensement serait spécialement utile lorsqu'il n'y a pas de programme de secteurs de recensement (SR). Une RD qui respecte les limites des secteurs de recensement accroît l'utilité du programme des secteurs de recensement en permettant d'approfondir davantage l'analyse des distributions à l'intérieur des SR. Par conséquent, après avoir examiné les différentes solutions de rechange qui représentaient des compromis entre la stabilité et l'utilité des RD, l'approche choisie devra

---

[3] Afin de regrouper les données de recensement en fonction des régions géographiques normalisées aux fins de diffusion, les SD doivent respecter les limites des régions géographiques normalisées. Ces limites sont portées à être modifiées d'un recensement à l'autre. Par exemple, entre les recensements, les limites municipales changent lorsque des municipalités sont créées ou dissoutes, ou lorsqu'elles annexent des parties de municipalités adjacentes (les municipalités sont une composante importante de la région géographique normalisée de la subdivision de recensement (SDR)). Par le passé, nous pouvions nous attendre à ce que les limites de 20 % des 6 000 SDR changent d'un recensement à l'autre. La circonscription électorale fédérale (CÉF) est une autre région géographique normalisée que les SD doivent respecter. Quand de nouvelles ordonnances de représentation relatives à la CÉF entrent en vigueur, 20% des SD de recensement sont modifiés. La croissance démographique et l'augmentation du nombre de logements s'y rapportant, les changements au réseau routier et les modifications apportées aux spécifications de la collecte sur le terrain, telles que le nombre maximal de logements par SD, entraînent aussi des modifications au SD d'un recensement à l'autre.

exiger que la RD respecte les limites des subdivisions de recensement (SDR) et des secteurs de recensement (SR).

Quelles sont les incidences de cette approche sur la stabilité de la RD ? On apporte rarement des modifications aux limites des SR, et seulement lorsque cela est essentiel. Par conséquent, le respect des limites des SR a des répercussions minimales sur la stabilité de RD. Toutefois, les divisions futures des SR et la création de SR ou les modifications apportées aux limites des SR – d'ordinaire situés à la périphérie d'une région métropolitaine de recensement (RMR) ou d'une agglomération de recensement (AR) – entraîneront des modifications aux RD. Toutefois, ces changements sont localisés et, une fois qu'ils sont mis en œuvre, le secteur local redevient stable puisque la zone de changement pousse vers l'extérieur. On prévoit aussi un changement lorsqu'une SDR est incorporée ou lorsqu'une limite de SDR change à la suite d'une annexion. De nouveau, on s'attend à ce que le changement soit localisé et ponctuel. En général, lorsque la croissance se produit, les divisions des RD sont probables, comme dans le cas du programme des SR. En cas de détérioration, il faudrait envisager des regroupements, en particulier en ce qui a trait au critère du seuil de la population minimale. De légères modifications aux limites d'un SR ou d'une SDR engendreraient dans la plupart des cas des changements semblables à la RD concernée, alors que d'importants changements (par exemple, une agglomération de recensement qui est intégrée au programme des secteurs de recensement) pourraient résulter en un remaniement complet de l'agglomération concernée.

Le critère de stabilité temporelle présente deux aspects. Dans ce qui précède, nous avons traité de la partie orientée vers l'avant – l'établissement d'une base solide pour la RD en commençant par le Recensement de 2001 et en maximisant la stabilite pour le Recensement de 2006 et au-delà. Toutefois, qu'en est-il de l'aspect inverse, notamment la conservation de la comparabilité avec le SD du Recensement de 1996 ? Pendant les consultations, l'opinion relative à ce point en particulier était plus divisée que dans le cas de tous les autres critères. Un groupe souhaitait clairement que la comparabilité avec le SD du Recensement de 1996 soit un critère de conception alors qu'un autre groupe préférait l'abandon de ce bagage historique. En fin de compte, c'est l'opinion du dernier groupe qui a prévalu parce que la conservation de la comparabilité signifiait aussi la conservation de RD sous-optimales. Cela aurait aussi limité la création de nouvelles RD, créant ainsi un contexte qui favoriserait l'établissement d'autres RD sous-optimales.

2.2 Population minimale (réduction de la suppression de données sur des régions)

Les critères servant à la délimitation des secteurs de dénombrement pour la collecte du recensement vont à l'encontre des techniques utilisées pour protéger la confidentialité des réponses individuelles à des fins de diffusion. Une technique appelée « suppression de données sur des régions » entraîne l'élimination de toutes les données sur les caractéristiques du recensement pour les régions géographiques où les chiffres de population sont inférieurs à 40 (inférieurs à 250 dans le cas des données sur le revenu)[4].

L'application des règles de suppression de données sur les 49 361 secteurs de dénombrement (SD) pour le Recensement de 1996 signifiait l'élimination des données sur le revenu pour 13 085 SD (27%) comptant moins de 250 personnes, alors que les autres données sur les caractéristiques étaient supprimées pour 4 765 SD (10 %) comptant moins de 40 personnes. En fait, 2 278 (5 %) des 49 361 SD en 1996 avaient une population zéro. La nécessité de réduire de tels niveaux élevés de suppression de données sur des régions pour la plus petite région géographique normalisée ou la plus petite région de diffusion est axiomatique. Par le passé, cela était impossible en raison du double rôle du secteur de dénombrement. À présent, cela est en grande partie possible dans le cadre de la conception de la région de diffusion.

On a déterminé qu'une population minimale de 500 personnes est le chiffre cible pour la région de diffusion. Une pondération de pénalité est appliquée à la solution automatisée lorsque le chiffre de

[4] L'autre technique utilisée est l'arrondissement aléatoire. Dans le cadre de cette méthode, tous les chiffres comprenant des totaux, qui sont supérieurs à 10 sont arrondis de façon aléatoire vers le haut ou le bas à un multiple de « 5 », et les chiffres représentant 10 ou moins sont arrondis de façon aléatoire à « 0 » ou à « 10 ».

population pour la RD se situe au-dessous de la valeur cible. La valeur de la pondération augmente à mesure que la différence entre la solution et le chiffre cible s'accroît. Le taux auquel la pondération de pénalité augmente peut aussi être contrôlé afin qu'on puisse modifier les incidences des différents critères en fonction de leur importance évaluée par rapport à la solution de la RD. On a attribué les pondérations de pénalité les plus élevées au critère de la population minimale. Néanmoins, les RD ayant des chiffres de population inférieurs à 500 seront quand même délimitées. La solution de la RD qui est acceptée est celle qui a la plus faible pondération de pénalité totale, qui est l'agrégat des pondérations de pénalité pour tous les critères combinés relativement à toutes les RD situées à l'intérieur de la plus grande région de délimitation[5, 6]. Toutefois, comme le critère de la population minimale a la plus grande pondération de pénalité, dans la plupart des cas, la RD n'aura pas de chiffres de population considérablement inférieurs à la valeur cible, et il est certain que les RD ayant moins de 250 personnes seront rares et donneront lieu à deux mises en garde.

La première mise en garde est que les RD doivent respecter les limites des SDR et des SR (en raison du critère de la « stabilité temporelle » décrit plus haut). Par conséquent, le système automatisé produira des solutions pour les plus grandes zones de délimitation définies par l'intersection des SDR et des SR. Un certain nombre de SDR et de SR se situeront au-dessous de la population cible. Par exemple, l'utilisation des résultats du Recensement de 1996 a produit 2 329 SDR et 94 SR avec une population inférieure à 500. Le nombre de SDR et de SR comptant moins de 250 personnes est de 1 449 et 69 respectivement, et le nombre comptant moins de 40 personnes est de 637 et de 40 respectivement. De plus, des SR chevauchent des SDR, ce qui fait que des plus grandes régions de délimitation ont des chiffres de population inférieurs au chiffre cible de 500. Il y a 256 SR qui chevauchent les limites des SDR. Supposez qu'environ 50 000 RD soient créées pour le Recensement de 2001 et que tous les 256 SR chevauchants créent de plus grandes régions de délimitation ayant moins de 40 personnes. Nous pouvons alors nous attendre à ce qu'au moins 6 % des RD compteront moins de 500 personnes, 4 % compteront moins de 250 personnes, et 2 %, moins de 40 personnes. Cette situation n'est pas parfaite mais elle représentera une nette amélioration par rapport aux secteurs de dénombrement.

La deuxième mise en garde concerne la source des données sur la population qui seront utilisées dans la délimitation des RD. Nous aimerions utiliser les chiffres de population du Recensement de 2001 par îlot. Cependant, la date à laquelle un ensemble final de RD sera requis peut exclure l'utilisation des données du Recensement de 2001. Cela signifie que la délimitation des RD devra utiliser des estimations de la population par îlot. La méthode relative à cet aspect de la création de RD n'a pas fait l'objet de recherche suffisante afin de décrire le processus ou la qualité prévue. Néanmoins, nous envisageons de distribuer les chiffres relatifs aux SD de 1996 à l'ensemble des îlots correspondants en faisant un certain ajustement pour la croissance. Comme conséquence de l'erreur associée à ce processus, nous prévoyons des cas où les chiffres de population de RD seront inférieurs au nombre visé de 500. Mais nous prévoyons aussi que les estimations seront suffisamment solides pour empêcher la création de RD ayant un chiffre de population inférieur à 250.

2.3 Population maximale (uniformité)

L'objet de ce critère est de garder uniforme la taille de la population de la RD. Ce critère aura l'avantage de produire plus de RD à l'intérieur de la plus grande région de délimitation. Cela sera utile quand on tentera d'intégrer des RD dans d'autres régions géographiques présentant de l'intérêt. Sans critère

---

[5] La plus grande région de délimitation est la région de travail géographique pour laquelle on recherche une solution. En ce qui concerne la RD, elle est définie par l'intersection des subdivisions de recensement (SDR) et des secteurs de recensement (SR).

[6] Cette approche ne tient pas directement compte de l'étendue des cotes individuelles pour chaque RD, à l'intérieur de la plus grande région de délimitation. Par conséquent, l'utilisation de la solution globale avec la cote de la plus faible pénalité agrégée à l'intérieur de la plus grande région de délimitation pourrait produire une mauvaise RD tout en produisant des solutions optimales pour les bonnes RD à l'intérieur de la plus grande région de délimitation. Une approche de rechange permettrait de tenir compte des cotes relatives aux RD individuelles de sorte que la création d'une mauvaise RD soit empêchée, mais cela causerait des solutions de légèrement moins bonne qualité pour les autres RD à l'intérieur de la plus grande région de délimitation, ce qui signifie que la solution relative à la cote de la plus faible pénalité agrégée ne serait pas choisie. Il faut évaluer si la deuxième approche est meilleure que l'approche actuellement en cours de programmation pendant la phase expérimentale.

explicite relatif à la population maximale, la délimitation aurait tendance à aller vers des RD plus grandes et moins nombreuses à l'intérieur de la plus grande région de délimitation.

La population maximale visée est aussi de 500. Ce n'est pas aussi crucial que le critère minimal parce qu'il n'influence pas la disponibilité des données. Par conséquent, les cotes de pénalité attribuées à la population maximale cible sont inférieures à celles attribuées à la population minimale cible. Par conséquent, *la tendance globale sera de produire des RD plus grandes que la population cible de 500,* mais de limiter les solutions afin de ne pas créer des RD qui dépassent considérablement l'objectif. Comme avec le critère de la population minimale, l'utilisation des estimations de population influencera les résultats finals mais cela n'est pas considéré être très grave dans ce cas[7].

Pendant les délibérations concernant les critères de population minimale et maximale, une question connexe a été soulevée. Il s'agissait de la méthode de sélection des populations cibles et de points relatifs à la fiabilité des données. En général, comme la fiabilité des données diminue avec la taille de la cellule, l'objectif est d'essayer d'avoir des tailles de cellule supérieures à une certaine taille minimale. Par conséquent, du point de vue de la fiabilité des données, plus le seuil minimal sera élevé, le mieux cela vaudra. Toutefois, le fait d'augmenter le seuil minimal au-delà d'un certain niveau va à l'encontre des objectifs d'une région de diffusion géographique normalisée de base; par conséquent, une limite maximale s'impose. La conception des tableaux est aussi un élément crucial quand il s'agit de la fiabilité des données. En fin de compte, essayer de produire des RD ayant une taille uniforme (en termes de la taille de la population) a été considéré comme l'objectif parce que l'uniformité résultante faciliterait la conception de tableaux de sortie optimaux. On a considéré que l'objectif de 500 était, dans l'ensemble, un choix approprié en ce qui a trait à la suppression de données sur des régions et aux préoccupations liées à la fiabilité des données.

2.4 Limites intuitives (visibles)

Les secteurs de dénombrement devaient respecter toutes les limites des régions géographiques normalisées pour qu'on puisse regrouper les données sur ces régions en vue de leur diffusion. À l'occasion, ces régions géographiques normalisées avaient des limites « invisibles », c.-à-d. quelques-unes des limites ou toutes les limites ne suivaient aucune caractéristique visible sur le terrain (telle qu'une rue, une voie ferrée, un détail hydrographique, une ligne de haute tension, etc.). Par conséquent, la limite du SD avait aussi une composante invisible. En outre, en l'absence de caractéristiques visibles appropriées, les limites invisibles de SD étaient parfois délibérément créées afin de limiter la taille du SD en vue de tenir compte des spécifications de la charge de travail de collecte. Les limites de SD qui suivent des caractéristiques invisibles peuvent rendre difficile, pour les utilisateurs, la détermination de la couverture géographique d'un SD, et l'incertitude associée peut soulever des questions au sujet de la qualité de l'application des données. Les utilisateurs ont donc demandé que l'on utilise seulement des limites visibles et facilement reconnaissables afin de délimiter les régions géographiques normalisées – qui ont été en général appelées limites « intuitives ».

Comme les RD sont des agrégats d'îlots, le critère de la limite intuitive est, dans la plupart des cas, assuré parce que les îlots respectent les routes. Cependant, les îlots doivent également suivre les limites de régions géographiques normalisées choisies et toutes les limites supplémentaires requises par l'opération de collecte en vue de limiter les volumes de travail. Comme on l'a indiqué plus haut, quelques-unes de ces limites (ou au moins des sections) sont invisibles. Par conséquent, la tâche pour la création de RD consiste à appliquer d'importantes cotes de pénalité aux limites des îlots qui sont invisibles. De nouveau, comme dans le cas d'autres critères, cela n'empêche pas la création de RD ayant des limites invisibles (ou au moins des limites partiellement invisibles), mais cela limite considérablement la probabilité de tels résultats. La condition exceptionnelle à l'énoncé précédent est le fait que la RD respecte les limites de toutes les subdivisions de recensement (SDR) et secteurs de recensement (SR) et, par conséquent, aura des limites invisibles dans la mesure que les SDR et les SR en ont.

[7] Les RD ayant une faible population se produiront lorsque la distribution de la population des SD de 1996 par îlot est inexacte, et les RD ayant une grande population se produiront lorsque l'ajustement de la croissance est inexact.

2.5 Forme compacte

Comme avec les limites intuitives, les exigences propres aux SD ont entraîné la production de SD ayant une forme étrange. Par exemple, lorsque deux régions géographiques régionales se chevauchaient, produisant une longue bande étroite où les deux limites se chevauchaient, il fallait créer une SD. De même, des formes étranges de SD résultaient quand l'accessibilité représentait un problème. Des SD aux formes étranges peuvent créer des problèmes quand les utilisateurs tentent de créer un « ajustement optimal » pour leur domaine d'intérêt au moyen du regroupement des SD. Les utilisateurs ont donc déterminé qu'il est nécessaire de rendre les régions de diffusion aussi compactes que possible.

Afin de répondre à ce besoin, un indice de compacité est inclus dans la conception des RD. L'indice est calculé en comparant la superficie de la RD à son périmètre. Un carré a une valeur d'indice de un (1,0) alors qu'un polygone dont la largeur est le double de la longueur a un indice de compacité de deux (2,0). En général, lorsqu'un polygone est étiré de sorte qu'un côté est « n » fois plus long que l'autre côté, l'indice de compacité est égal à « n ». Les valeurs supérieures à un (1,0) reflètent moins de compacité à mesure que la forme de la RD diffère de celle d'un carré parfait.

De nouveau, l'inclusion de ce critère n'empêche pas ce qui pourrait être jugé comme des résultats inappropriés. Lorsque ce critère est inclus, certains types de délimitations de RD deviennent seulement plus probables.

2.6 Homogénéité

Sur la base des soumissions présentées par les utilisateurs, l'homogénéité est le dernier facteur important dont il faut tenir compte pendant la délimitation d'une région de diffusion normalisée. Le type de logement est le plus souvent proposé comme la variable qui devrait être utilisée pour atteindre l'homogénéité voulue. Toutefois, des six facteurs clés dont l'inclusion dans la conception de la région de diffusion a été recommandée, c'est le seul que nous avons choisi de ne pas inclure explicitement dans la conception.

Cette décision est attribuable à l'absence de données appropriées par îlot. Le soutien de ce critère exigerait que des estimations du nombre de logements, par type de logement, soient produites par îlot à partir des données du recensement du niveau des SD de 1996. Il a été conclu que la qualité des estimations relatives au type de logement qui pourraient être produites serait simplement inadéquate pour la tâche; aucun critère d'homogénéité explicite n'a donc été inclus[8].

3. FACTEURS OPÉRATIONNELS

Outre les recommandations des utilisateurs, il y a aussi des facteurs opérationnels devant faire partie de la conception des régions de diffusion. Le plus important de ces facteurs est la structure du code de RD.

3.1 Structure du code de la région de diffusion

Au début du processus de conception, on a déterminé qu'il était nécessaire d'établir une structure hiérarchique représentée par un code de huit chiffres avec une proximité géographique incorporée. Par le passé, ces besoins opérationnels étaient déterminés ou antérieurement satisfaits par la structure province / circonscription électorale fédérale / secteur de dénombrement (PR / CÉF / SD). La taille du code (huit chiffres) est une conséquence de la taille du code PR / CÉF / SD, notamment 2 + 3 + 3 = 8. Cette taille de code est nécessaire pour éviter de remanier les nombreux programmes qui doivent avoir accès à ce code pendant le traitement du recensement. La structure hiérarchique avec la proximité

[8] Quand il est devenu évident que le critère d'homogénéité ne pouvait pas être directement soutenu, on a estimé pouvoir quand même soutenir l'homogénéité indirectement en maintenant la RD petite et compacte. Autrement dit, si les RD se composaient de deux à quatre îlots, l'homogénéité serait obtenue indirectement parce que le type de logements a souvent tendance à être cohérent d'un îlot à l'autre sans transitions soudaines. Bien que ce point de vue n'ait pas été justifié, il est probable que des RD comprenant de deux à quatre îlots seraient bien au-dessous de la norme en ce qui concerne la taille de la population.

géographique est nécessaire pour faciliter la révision et l'imputation quand le dossier d'un donateur proche géographiquement est requis.

Toutes les exigences de code seront satisfaites au moyen d'une hiérarchie province / division de recensement / région de diffusion (PR / DR / RD). Le code maximal de huit chiffres est satisfait par PR (2) + DR (2) + RD (4). Non seulement le besoin en matière de structure hiérarchique est satisfait, mais aussi le nombre d'unités à chaque niveau est environ le même puisqu'il y a 301 CÉF et 288 DR. Comme dans le cas des SD, la proximité géographique sera incorporée au code par l'attribution de codes d'une façon ondulante à l'intérieur des DR[9].

3.2 Code d'îlot à deux chiffres maximum

Un deuxième besoin opérationnel peut limiter le nombre d'îlots pouvant être inclus dans une RD à 99. Cela résulte d'une conception de la base de données qui fournit un code d'îlot à deux chiffres. Cette limite de la conception pourrait être modifiée (p. ex. : on pourrait créer un code à trois chiffres), mais il serait coûteux de le mettre en œuvre. Par conséquent, on a décidé d'attendre le premier essai de RD. Le premier essai ne comprendra pas cette limite. On prendra donc une décision définitive après avoir déterminé combien de fois des RD créées dépasseront la limite de 99 îlots et après avoir évalué les incidences que cette limite aura sur la création des régions de diffusion[10, 11].

3.3 Calendrier de production

On prévoit une période de production de temps écoulé de trois mois pour la création de la structure de la RD de 2001 (de mai à juillet 2001). Non seulement ce besoin opérationnel oblige à utiliser des estimations de population d'îlot pour la création de RD, mais en plus il peut influencer davantage les spécifications relatives aux RD si la durée de production requise dépasse la durée disponible. Cependant, un certain nombre de solutions éventuelles peuvent faire l'objet de recherche avant qu'il ne soit nécessaire de réduire les spécifications afin d'obtenir la réduction de la durée requise. D'abord, le pouvoir de calcul accru est une solution possible. La simplification des algorithmes sans réduire les besoins en spécifications est aussi une solution possible. Limiter le nombre de solutions à envisager en limitant la longueur de la durée de calcul pour une plus grande région de délimitation est possible, même si cette approche ne permet pas de trouver la meilleure solution. Il est important d'indiquer toutefois que la discussion de cette exigence est hautement spéculative et ne repose sur aucune expérience de production empirique à l'heure actuelle. Néanmoins, il s'agit d'un besoin opérationnel susceptible d'influencer les spécifications des RD; il est donc inclus dans le présent document par souci d'intégralité.

4. RÉTROACTION

Les spécifications relatives à la production de la région de diffusion pour le Recensement de 2001 sont en ce moment établies en grande partie. Toutefois, nous sommes toujours intéressés à recevoir de la rétroaction. Nous accueillons les commentaires, les suggestions ou les demandes de renseignements au sujet de tout aspect couvert par le présent document de travail. Ils devraient être adressés à Info-GÉO à la Division de la géographie.

REMERCIEMENTS

Le présent document de travail présente une compilation des idées et des recherches qui ont été développées au cours des cinq dernières années, et il est le fruit des efforts d'un certain nombre

[9] Cette structure de code est une exigence opérationnelle interne. Les structures de rechange à des fins de diffusion sont possibles et peuvent être examinées, c.-à-d. que nous encourageons les suggestions.

[10] Compte tenu de l'objectif de la population maximale de 500, nous estimons que le nombre de fois qu'une RD optimale dépasserait 99 îlots sera faible. Autrement dit, nous ne croyons pas que cette limite du nombre d'îlots dans une RD influencera négativement leur formation.

[11] La limite de 99 îlots s'applique aussi au nombre d'îlots par secteur de dénombrement. Par conséquent, une certaine élimination d'îlots s'ensuivra, mais les répercussions seront faibles.

d'employés de la Division de la géographie (Pessoa, Earl et Puderer, 1996). Les chercheurs suivants y ont contribué : Louise Earl, Peter Murphy et José Pessoa, ainsi que Carolyn Weiss et l'équipe de travail sur les régions de diffusion présidé par Rick Rideout qui ont fait les revues. Sincères remerciements à Karole Kidd qui a corrigé la version finale de ce document de travail et à Réjeanne Loranger qui a révisé la traduction du document.

RÉFÉRENCES

Pessoa, José, Earl, Louise et Puderer, Henry (1996). *The 2001 Logical Dissemination Unit Model: A Proposal for the Development of a New Small Area Geographic Unit for the Dissemination of Data*, rapport non publié, Division de la géographie (Ottawa : Statistique Canada).

Statistique Canada (1994). *Rapport sur les consultations en vu du Recensement de 1996*, Recensement du Canada de 1996, Rapport n° 2 (Ottawa : Statistique Canada).

Statistique Canada (1999). *Rapport sur les consultations en vue du Recensement de 2001-Géographie*, Recensement du Canada de 2001, N° 92-126-SIB au catalogue (Ottawa : Industrie Canada).

Annexe 1

L'ÎLOT

Jusqu'au recensement de 1996, les logements et leurs occupants (c.-à-d. les données du recensement) étaient géographiquement rattachés au secteur de dénombrement au moment de la collecte. À compter du Recensement de 2001, les données seront géographiquement rattachées à l'îlot au moment de la collecte[12].

Tout d'abord, les îlots sont des polygones formés par l'intersection de rues[13] (consulter les cartes 1 et 2). Ces îlots naturels ou morphologiques seront divisés afin de former deux îlots toutes les fois qu'un îlot est traversé par les limites de régions géographiques normalisées choisies, notamment les circonscriptions électorales fédérales (CÉF), les subdivisions de recensement (SDR), les secteurs de recensement (SR) ou les localités désignées (LD). Cette mesure est nécessaire pour qu'on soit capable de regrouper les données en fonction de toutes les régions de diffusion géographiques normalisées aux fins de diffusion.

Un îlot peut aussi être formé pour satisfaire à des besoins de collecte. Par exemple, les limites des volumes de travail exigent que des îlots soient formés pour les grands immeubles à appartements ou les immeubles collectifs. Dans ces cas, il en résultera des îlots à l'intérieur d'îlots. De plus, lorsque le réseau routier est clairsemé ou même inexistant, les limites des secteurs de dénombrement peuvent être établies afin de réduire le secteur de couverture à l'intention des recenseurs. Dans ces situations, l'îlot respectera la limite du secteur de dénombrement.

Les îlots sont regroupés pour établir des secteurs de dénombrement et des régions de diffusion. Comme les îlots sont principalement un produit découlant du réseau routier, le nombre d'îlots et leur création dépendent de la mesure dans laquelle la base de données sur le réseau routier sera à jour avant le recensement. Il n'est pas possible d'avoir un réseau routier qui reflète exactement la situation le jour du recensement. Cette situation se produira spécialement dans les régions de forte croissance[14].

Les chiffres de population et le nombre de logements sont diffusés par îlot. Les données sur les caractéristiques ne seront pas disponibles par îlot. La plus petite région géographique normalisée pour laquelle des données sur les caractéristiques seront fournies sera la RD. Toutefois, l'îlot remplacera le SD à l'extérieur de la couverture du côté d'îlot comme unité de base pour l'établissement de secteurs définis par les utilisateurs, fournissant ainsi la précision accrue dans la délimitation de ces secteurs.

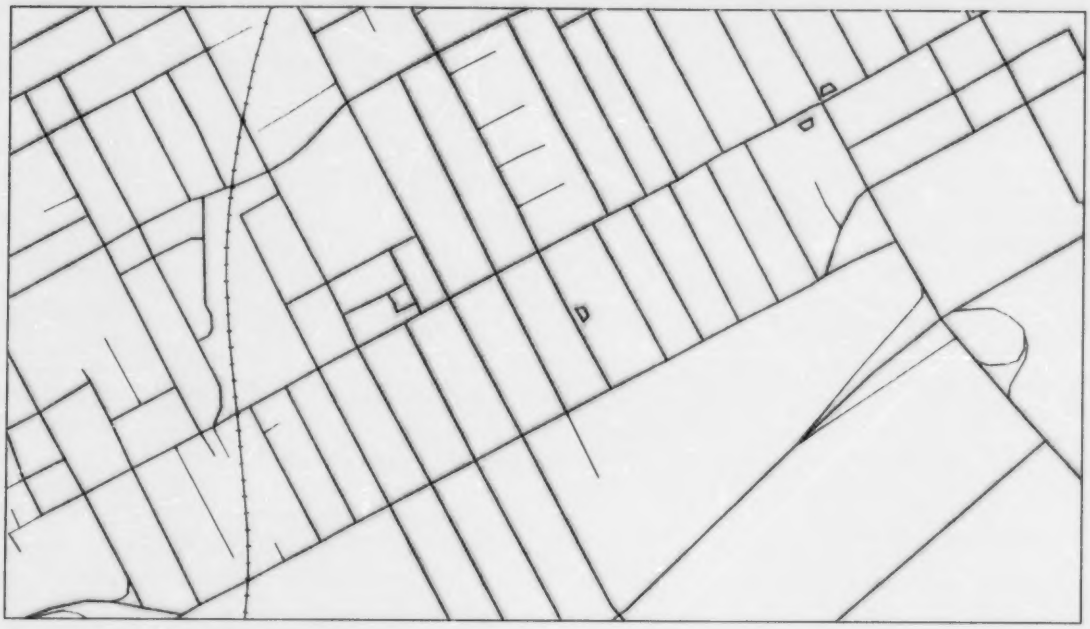
---

[12] Les données de recensement sont et continueront à être liées au côté d'îlot afin de faciliter l'extraction des données pour des secteurs définis par les utilisateurs. Cependant, ce lien au côté d'îlot est effectué comme activité suivant la collecte et seulement pour des secteurs choisis.

[13] Certains îlots morphologiques (c.-à-d. les terre-pleins, les bretelles d'accès et de sortie) seront fusionnés à des îlots adjacents.

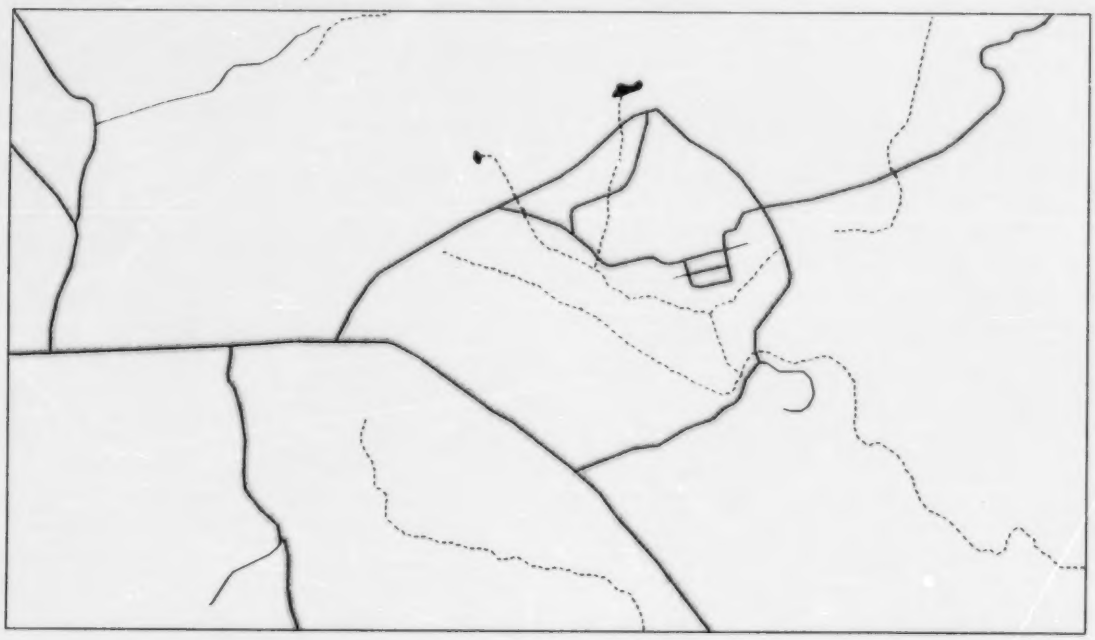
[14] Au moment de la rédaction du présent document, le réseau routier sur la BGN était à jour jusqu'à environ 1995 et a produit quelque 630 000 îlots. Toutefois, on continue à le mettre à jour. La dernière série de mises à jour en cours mettra la base à jour à compter du printemps de 1997.

Carte 1. Îlots dans une région urbaine pour le Recensement de 2001



Îlots
Routes
Chemins de fer
Lacs / étangs
Rivières / cours d'eau

Carte 2. Îlots dans une région rurale pour le Recensement 2001



Série de documents de travail de la géographie

Numéro

1993-1 ***Une comparaison des régions géographiques de recensement au Canada et aux États-Unis***, Carolyn Weiss, Michael Ratcliffe et Nancy Torrieri (novembre 1993)

1996-1 ***Qu'adviendra-t-il du calcul de la superficie des terres ? Justification du maintien du statu quo pour le recensement de 1996***, Carolyn Weiss (décembre 1996)

1998-1 ***Tests automatisés de la superficie des terres pour le recensement de 2001 : résultats préliminaires obtenus en utilisant les fichiers numériques cartographiques de 1996***, Carolyn Weiss et Augustine Akuoko-Asibey (avril 1998)

2000-1 ***Zones d'influence des régions métropolitaines de recensement et des agglomérations de recensement (ZIM) accompagnées de données du recensement***, Sheila Rambeau et Kathleen Todd (janvier 2000)

2000-2 ***Zones d'influence des régions métropolitaines de recensement et des agglomérations de recensement (ZIM) : Description de la méthodologie***, Chuck McNiven, Henry Puderer et Darryl Janes (janvier 2000)

2000-3 ***Délimitation du Nord canadien : un examen de la relation nord-sud au Canada***, Chuck McNiven et Henry Puderer (janvier 2000)

2000-4 ***Présentation de la région de diffusion pour le Recensement de 2001***, Henry Puderer (mars 2000)